Reliability Growth Planning Based on Essential Function Failures

Jonathan L. Bell, PhD, Institute for Defense Analyses

Steven D. Bearden, Jr., PhD, Institute for Defense Analyses

Key Words: Reliability, Reliability Growth Planning, PM2, Essential Function Failures

SUMMARY & CONCLUSIONS

The Department of Defense (DoD) has recently implemented a number of policies aimed at improving weapon system reliability and reducing long-term operating and sustainment costs. Military systems under DoD oversight are now required to have a reliability growth plan that describes how they intend to reach reliability goals during system development. Part of planning involves the construction of reliability growth planning curves to portray how a system's reliability is expected to increase over time.

Many DoD programs base their reliability growth planning curves on a mission-level reliability requirement that includes only failures discovered during mission execution that result in an abort or termination of a mission in progress. However, the majority of failures that occur during testing do not lead to mission aborts, such as those discovered outside of the mission window or failures of redundant or non-mission essential components. Reliability metrics that are limited to mission aborts are an important measure of mission effectiveness but are not well suited for defining a robust reliability growth strategy because, in many cases, they exclude a large portion of the failure modes that drive maintenance costs and reduce system availability.

An alternative approach would be to base requirements and reliability growth curves on a measure that includes all failures of mission essential components that drive maintenance costs and degrade system availability, regardless of when the failure is discovered. Using examples and lessons learned from a recent military aircraft program, this article describes why this approach produces a more credible and effective reliability growth strategy in a less resource-intensive way by:

- Incorporating a larger share of the failure modes
- Addressing problems before they turn into mission aborts
- Improving the ability to assess and track reliability growth
- Increasing the statistical power and confidence to evaluate reliability in testing
- Enabling more reasonable reliability growth goals
- Reducing subjectivity that can creep into the reliability scoring process
- Correcting a larger share of the failures that negatively affect system availability, maintainability, and operating and sustainment cost.

1 INTRODUCTION

1.1 *Reliability Requirements in the DoD*

Reliability requirements in the DoD are often categorized as mission-level or logistic-level requirements. Table 1 describes commonly used reliability measures and their associated definitions. Logistic metrics such as MTBUMA and MTBF typically encompass all deferrable and nondeferrable failures of the system. Essential function failures or essential maintenance actions are a subset of failures that result in the degradation or loss of a mission essential capability. A further subclassification concerns failures of essential subsystems that occur during a mission resulting in a termination of the mission in progress. These failures are commonly categorized as mission aborts, mission failures, or operational mission failures. Depending on the system type, the time aspect of these measures could be stated in terms of flight hours, miles, or rounds.

	Reliability Measure[a]	Definition
Logistic-Level Requirements	Mean Time Between Unscheduled Maintenance Actions (MTBUMA)	Includes all failures or required unscheduled maintenance actions for the system, regardless of the time of discovery
	Mean Time Between Failures (MTBF)	
	Mean Time Between Essential Function Failures (MTBEFF)	Incidents or malfunctions discovered anytime that cause the inability to perform one or more mission essential functions; usually result in the system being declared non-operational
	Mean Time Between Essential Maintenance Actions (MTBEMA)	
Mission-Level Requirements	Mean Time Between Mission Aborts (MTBMA)	Incidents or malfunctions that occur during a mission or before the start of a mission resulting in the loss of an essential function required for the mission in progress; as a result, the user is required to terminate the mission
	Mean Time Between Mission Failures (MTBMF)	
	Mean Time Between Operational Mission Failures (MTBOMF)	

Table 1 – Common DoD Reliability Requirements

Generally, the number of mission terminating failures is small relative to the overall number of failures observed during testing. Non-mission terminating failures would include failures discovered outside of the mission window or failures that occur during missions that do not cause mission aborts. All mission aborts would also be counted as essential function failures (EFF). Similarly, all EFFs would also be counted toward the total number of failures.

Most DoD systems establish and define their primary measure of reliability using a mission-level metric such as MTBMA. As a result, the reliability growth program and planning curves are also based on a mission-level measure of reliability. For example, Figure 3 illustrates a reliability growth planning curve using the Planning Model based on Projection Methodology (PM2), along with the associated model parameters [1-2]. Since its development, the PM2 model has been widely used within the DoD for reliability growth planning and has been specifically required for Army programs since June 2009.

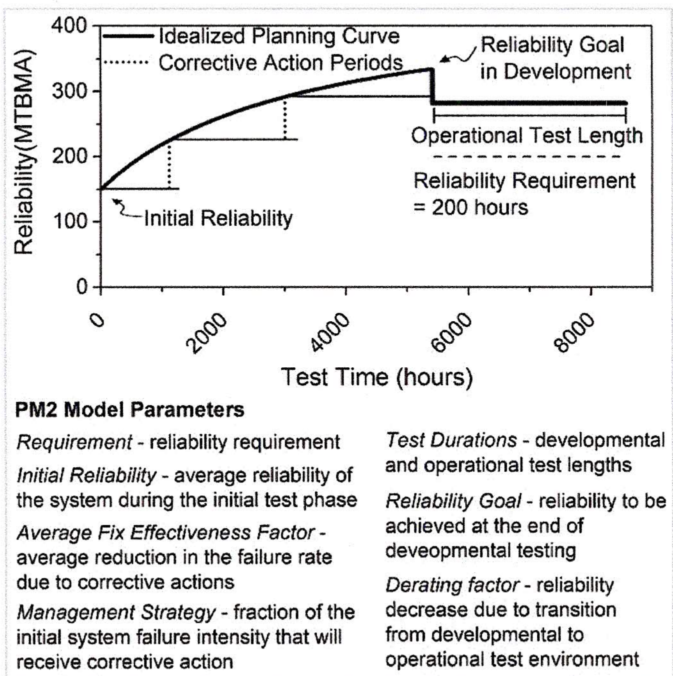

Figure 1 – PM2 Reliability Growth Planning Model

The PM2 model permits scheduling specific blocks in time, shown as flat steps under the idealized reliability growth curve, known as corrective action periods, when corrective action would be implemented. The reliability goal in PM2 is usually based on demonstrating the reliability requirement with a selected level of producer and consumer risk during operational testing. Therefore, the reliability goal is larger than the requirement to ensure that the requirement is demonstrated during operational testing with an acceptable level of statistical confidence and power. In setting the reliability growth goal, the model also provides the option to use a derating factor to account for the drop in reliability that is often observed as a system transitions from a developmental to operational test environment.

1.2 *Poor Reliability Trends in the DoD*

Data from reports published by the Director, Operational Test and Evaluation (DOT&E) confirm that many weapon systems fail to demonstrate their established reliability requirements during operational testing. Figure 2 illustrates data for systems that had an operational test between 1997 and 2013. In the 124 reports published during this time period, DOT&E concluded that 108 systems were effective and only 67 were reliable. The bar chart at the bottom of Figure 2 shows that the fraction of systems that met their reliability requirements in each year is often less than 50 percent.

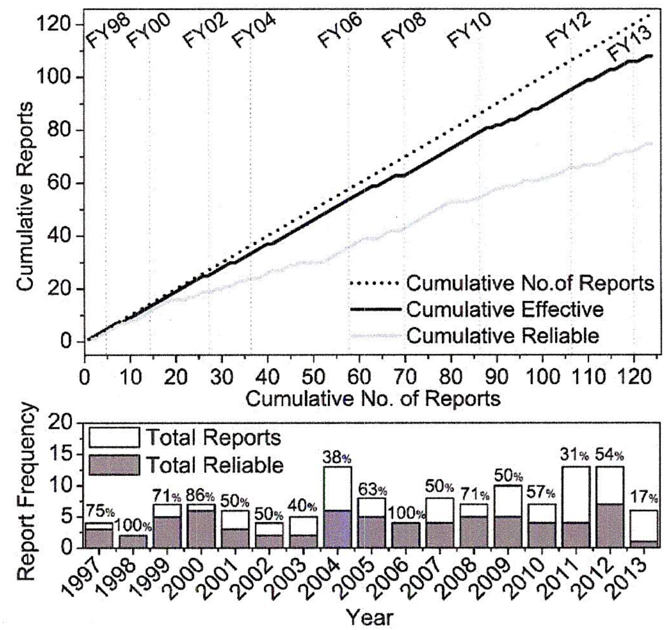

Figure 2 - Summary of DOT&E Reports

In 2008, a reliability improvement Task Force was set up within the Defense Science Board to investigate the poor reliability trends. The Task Force concluded that the most important step in correcting high suitability failure rates is to ensure programs are formulated to execute a viable systems engineering strategy from the beginning, including a robust reliability, availability, maintainability program, as an integral part of design and development [3].

Following this investigation, the DoD instituted policies aimed at improving weapon system reliability and reducing long-term operating and sustainment costs. Policies now require weapon systems to have a reliability growth program as part of system design and development, and reliability growth management that is fully integrated across systems engineering, life cycle sustainment, and test and evaluation activities. Since November 2009, programs under DOT&E oversight have also been required to include a reliability growth planning curve within their test planning documentation.

Despite recent DoD guidance, many systems still fail to reach reliability objectives. Part of the problem is that reliability requirements and reliability growth planning curves are often based on a mission-level reliability requirement.

This approach is not well suited for building a robust reliability growth strategy, because it discounts failure modes that are discovered outside of the mission execution window and excludes many of the failure modes that drive maintenance costs and reduce system availability. In fact, mission-based reliability metrics consider only a subset of the overall system reliability and are more a measure of mission effectiveness than an all-inclusive measure of reliability.

Using examples and lessons learned from a military aircraft program, this article discusses the advantages of basing growth planning curves on a requirement that includes all mission essential failures of the system.

2 *ADVANTAGES OF RELIABILITY GROWTH CURVES BASED ON ESSENTIAL FUNCTION FAILURES*

2.1 *Incorporation of a Larger Share of the Failure Modes*

EFF-based reliability growth planning curves capture a larger portion of the relevant failure modes, because a program will generally observe more EFFs than mission aborts. The aircraft program considered in this article (called "System-A" henceforth) had 422 EFFs, and since 68 of the 422 EFFs were discovered during missions leading to a loss of mission capability, they were classified as mission aborts. Some EFFs occurred during missions without causing mission aborts because the failures did not degrade a mission essential function or there was a redundant capability that enabled mission completion. For example, military aircraft have more than one radio; failure of a single radio would not necessarily lead to a mission abort. The 422 EFFs comprised failures of mission essential components such as main transmission seals, aircraft survivability equipment, and the hydraulics systems.

For System A, numerous failure modes were responsible for the 422 EFFs; far fewer failure modes were responsible for the 68 mission aborts. Reliability growth planning curves based on mission aborts would be limited to failure modes discovered during missions. This would leave open the potential for uncorrected failure modes to follow the system into production and fielding.

Although the primary reliability requirement for most DoD system is a mission-level metric, System-A saw the value in closely tracking all EFFs and decided to construct two reliability growth planning curves: one based on its mission abort requirement and a second based on its EFF requirement.

2.2 *Correction of Failure Modes Before they Turn into Aborts*

Correcting EFFs discovered in early developmental testing reduces the risk that these failures could reoccur during missions and cause a mission abort. For example, System-A added new mission processors to reduce aircraft wiring and weight. Early in its development, the system experienced a number of EFFs due to mission processor failures. The producer determined that more than one failure mode was responsible for the mission processor failures and implemented hardware and software corrective actions to address the problems. During its recent operational test, mission processor failures did not occur during any of the missions. If both mission processors had failed during a mission, it would have caused a mission abort.

2.3 *Improved Ability to Assess and Track Reliability Growth*

After a program begins collecting reliability data, it is necessary to track the progress of reliability efforts and, in some cases, to project what reliability could be attained at a future time. If reliability data are limited, it will be difficult to perform these assessments. For example, Figure 3 illustrates the cumulative number of EFFs and mission aborts for a System-A prototype during the first 100 flight hours of testing along with a fit to the data using the AMSAA-Crow reliability growth tracking model. There were 25 EFFs and 3 mission aborts during this time. The AMSAA-Crow model fit for cumulative EFFs is reasonably good, indicating positive reliability growth with a growth rate (α) of 0.17. However, due to the limited number of data points, the fit produced for mission aborts is subject to more uncertainty and suggests a growth rate of -2.24.

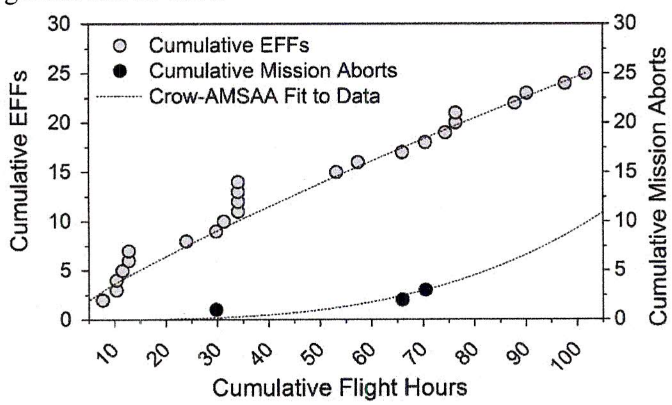

Figure 3 – Cumulative EFFs and Mission Aborts for System-A

A program might also wish to project future reliability following the implementation of corrective actions. For example, the System-A program held reliability assessment conferences to estimate the Fix Effectiveness Factors of corrective actions. Considering the data in Figure 3, if System-A implemented corrective actions only for mission aborts, the evaluation of fix effectiveness would be limited to failure modes causing the three aborts. The greater number of failure modes related to EFFs provides a much better estimate of the average Fix Effectiveness Factor, particularly when it is early in the program and the data are limited.

2.4 *Greater Statistical Power and Confidence to Evaluate Reliability During Operational Testing*

In planning operational tests, it is desirable to test the system long enough under operationally realistic conditions to be able to estimate its reliability with statistical confidence. A common approach is to use operating characteristic curves to determine the test length necessary to demonstrate the system's reliability requirement with acceptable levels of consumer and producer risks. In many instances, the reliability requirement based on mission aborts is large and, as a result, lengthy testing is needed. Requirements based on

EFFs are typically 5 to 10 times smaller than those for mission aborts and therefore require shorter testing to estimate the EFF requirement with statistical confidence.

Figure 4 illustrates operating characteristic curves for System-A's MTBEFF and MTBMA requirements. The operational test length is on the left vertical axes with the corresponding number of mission aborts or EFFs permitted on the right vertical axes, the true system reliability is on the x-axes, and the color coding quantifies the producer risk. The consumer risk is set at 0.2. For MTBEFF, the minimally acceptable MTBEFF requirement is 2.3 hours with a MTBEFF system specification of 3.1 hours. For MTBMA, System-A's minimally acceptable reliability requirement is 15.3 hours with an MTBMA system specification value of 22 hours.

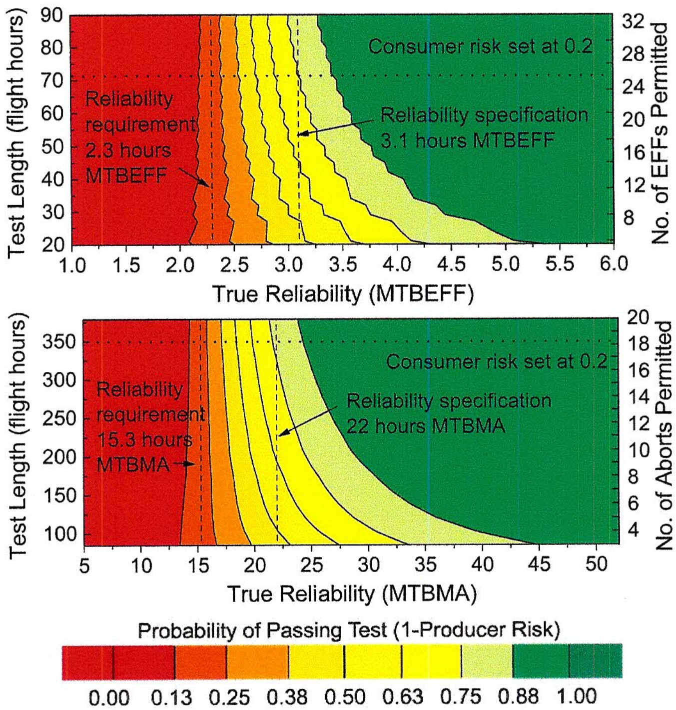

Figure 4 – System-A Operating Characteristic Curves for MTBEFF (top) and MTBMA (bottom)

In general, systems should strive to test long enough to evaluate mission-level reliability requirements. However, in situations where the mission-level requirement is too large to evaluate within a reasonable timeframe, testing an EFF-based requirement can provide a reasonable alternative. For System-A, the reliability requirement for MTBEFF is smaller than the MTBMA requirement (2.3 hours compared to 15.3 hours for MTBMA), so fewer test hours are necessary to demonstrate the MTBEFF requirement with acceptable consumer and producer risk. For example, if the producer delivers an aircraft with a true reliability of 22 hours MTBMA, a 350-hour test provides a 75 percent probability of demonstrating the 15.3-hour MTBMA requirement with 18 or less aborts. In comparison, if the producer delivers an aircraft with a true reliability of 3.1 hours MTBEFF, only 72 hours of testing are required to have a 75 percent probability of demonstrating the 2.3-hour MTBEFF requirement with 26 or less EFFs.

2.5 *Development of More Reasonable Reliability Growth Goals*

For systems with a large mission-level reliability requirement, it is challenging to construct an executable growth planning curve because either a very long test is required or the growth goal in development is unreasonably high. For example, considering the MTBMA operating characteristic curves in Figure 4, if System-A was constrained to a 100-hour test, the program would need to grow the true system reliability in development to approximately 40 hours MTBMA to have an 80 percent chance of demonstrating 15.3 hours MTBMA with 80 percent confidence. An MTBMA of 40 hours is well above the system specification of 22 hours MTBMA. If the program is only funded to grow reliability to the system specification of 22 hours MTBMA, a 659-hour test is required to have an 80 percent chance of demonstrating a 15.3-hour MTBMA with 80 percent confidence.

In such situations, growth curves based on EFFs allow for more reasonable growth goals and test lengths. The corresponding System-A MTBEFF growth goal for a 100-hour test is 3.0 hours MTBEFF, which is below the system specification of 3.1 hours. In other words, if System-A's true reliability is 3.0 hours MTBEFF, it would have an 80 percent chance of passing a 100-hour test by demonstrating 2.3 hours MTBMA with 80 percent confidence.

2.6 *Reduced Subjectivity During Reliability Scoring*

EFF-based reliability measures are more applicable and less subjective than mission-abort based measures, particularly during developmental testing (DT). For most programs, the majority of the reliability growth program is conducted during DT. System-A accrued two-thirds of its flight hours in DT and one-third in operational testing (OT). Scoring of mission aborts in DT was often difficult and somewhat artificial because:

- DT missions are typically not time sensitive; the mission begins when the system is ready. In OT, mission start times are more strictly enforced to better reflect the operational environment.
- Systems in DT are not always fully equipped with all radios, weapons, survivability equipment, and other devices, any of which can contribute to mission aborts. Failures of subsystems not being tested in DT might not be scored as in the same way as they would be during OT, because there can sometimes be differing views on whether a particular subsystem that had a failure was required for the mission in progress.
- Civilian contractors rather than soldiers maintain the system during DT and assist in preparing the system for mission execution. As a result, maintenance-induced failures leading to an abort are less likely to occur in DT.
- Developmental testers are usually more experienced than soldiers and are less likely to use the system in a way that causes a mission abort. Therefore, the program is less likely to observe user-induced aborts in DT.

EFFs are less sensitive to these concerns because they are counted whether they were discovered during a mission or not, irrespective of mission timing or whether the capability was specifically required for the mission in progress.

2.7 *Correction of a Larger Share of the Failures that Negatively Affect System Availability, Maintainability, and Cost*

A reliability growth strategy based on EFFs can result in larger improvements in system availability, maintainability, and life cycle costs. As defined for System-A, EFFs are failures that cause system downtime. System availability will be improved by correcting more of the failure modes that cause downtime. An EFF-based reliability growth plan with a robust management strategy will correct a larger portion of the failures that require essential maintenance. As failure frequency is reduced, maintenance demand decreases, fewer replacement parts are required, and ultimately, operating and sustainment costs are reduced. Because the program for System-A strived to correct EFF-based failure modes during development, its demonstrated reliability and maintainability were far better than required during operational testing.

3 CONCLUSIONS AND RECOMMENDATIONS

Compared to reliability growth planning curves based on a mission-level reliability requirement, EFF-based reliability growth plans are more likely to identify and correct failure modes that cause system downtime, leading to greater improvements in reliability, availability, and maintainability, and reduced operating and sustainment costs. EFF-based curves also provide more failure data early in the program, which improves the ability to track reliability growth and the implementation of corrective actions and to estimate reliability during operational testing. This approach is more consistent with the DoD reliability Task Force recommendation of ensuring that programs include a robust Reliability, Availability, Maintainability program, as an integral part of development.

In the future, programs should establish EFF-based reliability requirements and base reliability growth planning curves on this requirement. In some situations, it might be useful to construct two reliability growth planning curves: one based on an EFF requirement and another based on the system's primary reliability requirement, which is usually defined at the mission abort level. Doing this provides the program with the flexibility to tailor each curve as needed for the system. For programs that have a large mission abort-based reliability requirement, growth planning based on EFFs is particularly attractive because it is often difficult to develop an executable growth planning curve or conduct an operational test that is long enough to assess the mission abort requirement.

REFERENCES

1. P. M. Ellner, J. B. Hall, "An Approach to Reliability Growth Planning based on Failure Mode Discovery and Correction," *Proc. Ann. Reliability & Maintainability Symp.,* (Jan.) 2006, pp 266-272.
2. J. B. Hall, "Reliability Growth Planning for Discrete-Use Systems," *Proc. Ann. Reliability & Maintainability Symp.,* (Jan.) 2011, pp 1-6.
3. "Report of the Defense Science Board Task Force on Developmental Test & Evaluation," U.S. Department of Defense, Washington, D.C., May 2008.
4. L. H. Crow, "Reliability Analysis for Complex, Repairable Systems in Reliability and Biometry," SIAM, ed. by Proschan and R. J. Serfling, Philadelphia, Pennsylvania, 1974, pp 379-410.

BIOGRAPHIES

Jonathan L. Bell, PhD
Institute for Defense Analyses
Operational Evaluation Division
4850 Mark Center Drive
Alexandria, VA 22310 USA

e-mail: jlbell@ida.org

Jonathan L. Bell is a Research Staff Member at the Institute for Defense Analyses (IDA). He earned his doctoral degree in Materials Science and Engineering at the University of Illinois at Urbana Champaign in 2008 and his bachelors degree in Materials Science at Carnegie Mellon University. Dr. Bell's work at IDA is focused on operational and live fire testing of rotary wing aircraft ground vehicle systems with a specific emphasis on reliability.

Steven D. Bearden, Jr., PhD
Institute for Defense Analyses
Operational Evaluation Division
4850 Mark Center Drive
Alexandria, VA 22310 USA

e-mail: sbearden@ida.org

Steven D. Bearden, Jr. is a Research Staff Member at the Institute for Defense Analyses. He earned his doctoral degree in Engineering with an emphasis in nanotechnology from Louisiana Tech University in Ruston, La. Dr. Bearden conducts research and analysis into the operational testing of land warfare command, control, and communication equipment including vehicular and soldier mounted networking radio systems.